

tRNADB-CE 2011: tRNA gene database curated manually by experts

Takashi Abe^{1,*}, Toshimichi Ikemura¹, Junichi Sugahara², Akio Kanai², Yasuo Ohara¹, Hiroshi Uehara¹, Makoto Kinouchi³, Shigehiko Kanaya⁴, Yuko Yamada¹, Akira Muto⁵ and Hachiro Inokuchi¹

¹Nagahama Institute of Bio-Science and Technology, Nagahama, Shiga, 526-0829, ²Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, 997-0035, ³Department of Bio-System Engineering, Graduate School of Science and Engineering, Yamagata University, Yonezawa, Yamagata, 992-0038, ⁴Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma, Nara, 630-0192 and ⁵Faculty of Agriculture and Life Science, Hirosaki University, Hirosaki, Aomori, 036-8561, Japan

Received September 10, 2010; Accepted October 7, 2010

ABSTRACT

We updated the tRNADB-CE by analyzing 939 complete and 1301 draft genomes of prokaryotes and eukaryotes, 171 complete virus genomes, 121 complete chloroplast genomes and approximately 230 million sequences obtained by metagenome analyses of 210 environmental samples. The 287 102 tRNA genes in total, and thus two times of the tRNA genes compiled previously, are compiled, in which sequence information, clover-leaf structure and results of sequence similarity and oligonucleotide-pattern search can be browsed. In order to pool collective knowledge with help from any experts in the tRNA research field, we included a column to which comments can be added on each tRNA gene. By compiling tRNAs of known prokaryotes with identical sequences, we found high phylogenetic preservation of tRNA sequences, especially at a phylum level. Furthermore, a large number of tRNAs obtained by metagenome analyses of environmental samples had sequences identical to those found in known prokaryotes. The identical sequence group, therefore, can be used as phylogenetic markers to clarify the microbial community structure of an ecosystem. The updated tRNADB-CE provided functions, with which users can obtain the phylotype-specific markers (e.g. genus-specific markers) by themselves and clarify microbial community structures of ecosystems in detail. tRNADB-CE can be accessed freely at <http://trna.nagahama-i-bio.ac.jp>.

INTRODUCTION

More than 99% of microorganisms that inhabit natural environments are difficult to culture under laboratory conditions. Metagenomic analyses of mixed genome samples have been developed (1–3) to explore the enormous number of novel genome resources. In accord with the remarkable progress of DNA-sequencing technology, a vast quantity of metagenomic sequences obtained from a wide variety of environmental samples have been decoded and released from DDBJ/EMBL/GenBank. Because a significant portion of environmental DNA sequences is derived from unculturable microbes, we can acquire new knowledge of tRNA sequences from novel genomes. The 154 455 tRNA genes found in the metagenomic sequences were included in the tRNADB-CE. When we focused on a group of tRNAs with an identical sequence, we found tRNAs found only in a particular lineage of phylogenetic groups. Notably, such phylotype-specific tRNA sequences were found in many species-unknown genomic fragments obtained by metagenome analyses of environmental samples. This shows that tRNA is a good phylogenetic marker for discovering phylotype composition and microbial community structure in an environmental ecosystem.

MATERIALS AND METHODS

The following sources of DNA sequences were used: the complete genomes of 927 prokaryotes and of 171 viruses released by Genome Information Broker (GIB, <http://gib.genes.nig.ac.jp/>) and Genome Information Broker for Viruses (GIB-V, <http://gib-v.genes.nig.ac.jp/>) of DDBJ up to March 2009; the complete genomes of 121 chloroplasts released by Organelle Genome Database (GOBASE, <http://gobase.bcm.umontreal.ca/>) up to March 2009; the

*To whom correspondence should be addressed. Tel: +81 749 64 8100; Fax: +81 749 64 8140; Email: takaabe@nagahama-i-bio.ac.jp

draft genome sequences of 1301 prokaryotes released by WGS division of DDBJ/EMBL/GenBank up to August 2009; the complete genomes of 12 eukaryotes and the 17 million metagenomic sequences released by DDBJ/EMBL/GenBank up to March 2010; and the 217 million metagenomic sequences obtained using a next-generation sequencer and released by Sequence Read Achieve (SRA, <http://www.ncbi.nlm.nih.gov/Traces/sra/>) in NCBI up to March 2010.

RESULTS AND DISCUSSION

Update of registered tRNA genes and a new function for organizing collected knowledge

The 287 102 tRNA genes in total (53 936, 70 079, 961, 3534, 4137 and 154 455 genes from 927 complete prokaryote genomes, 1301 draft prokaryote genomes, 171 complete virus, 12 complete eukaryote genomes, 121 complete

chloroplast genomes and 210 metagenomic samples, respectively) were registered in the updated tRNADB-CE. This was two times as many tRNA genes as registered in the previous version (4). This exhaustive search for tRNA genes was performed by running three computer programs used for tRNA gene search, tRNAscan-SE (5), ARAGORN (6) and tRNAsfider (7) in combination. This method should enhance the completeness and accuracy of prediction since the programs' algorithms are partially different and render somewhat different results. The tRNA genes found accordantly by all three programs were stored in tRNADB-CE without further checks. Then, the residual, discordant cases (~4% of the total of tRNA gene candidates), except for those found in metagenomic sequences, were checked, manually by three experts (Y.Y., A.M. and H.I.) in the tRNA experimental field and were classified into three categories: (i) reliable tRNA genes, (ii) not tRNA genes and (iii) ambiguous cases. Users can browse or download

tRNADB-CE: tRNA gene database curated manually by experts
 Takeshi Abe¹, Toshiyuki Ikenawa¹, Junichi Sugahara², Atsushi Kondo³, Yasuo Ohara⁴, Hiroshi Ueda⁵, Makoto Kinouchi¹,
 Shigenobu Konno¹, Yuji Yoneda¹, Akira Muto², Masahiro Inouchi¹
¹ Nagoya Inst. of Bio-Sci. and Tech., ² Keio Univ., ³ Yamagata Univ., ⁴ NIST, ⁵ Hiroshima Univ.
[Top](#) | [Keyword](#) | [BLAST](#) | [Download](#) | [HOW TO USE](#)

Data List

Data Type	Genomes	tRNA genes
Prokaryote	Bacteria	880 51796
	Archaea	47 2140
Draft	1301 70079	
Eukaryote	Plant	2 1346
	Fungi	10 2188
Virus		31 151
Phage		140 810
Chloroplast		121 4137
Environmental sample (ENV)		123 99683
ENV from Sequence Read Achieve (SRA)		87 54872

Keyword Search
 Keyword: (Enter a keyword separated with a space.)
 Targets: (Select target species)
 Data Type: (Select data type)
[Advanced Search](#)

Sequence Search (BLASTN/Advanced Pattern Search)
 Query: (Enter query sequence)
 MURF: (Checkmark)
 File Type: (FASTA, Tab delimited data)
 Data Types: (Bacteria, Archaea, Draft, Plant, Fungi, Virus, Phage, Chloroplast, ENV, SRA)
 Target: (Enter target sequence)
[Advanced Pattern Search](#)

tRNA gene data download
 File Type: (FASTA, Tab delimited data)
 Data Types: (Bacteria, Archaea, Draft, Plant, Fungi, Virus, Phage, Chloroplast, ENV, SRA)
 Target: (Enter target sequence)

(a) Phylum/Class/species list and anticodon table

Phylum/Class/Species list

Search species: [Search](#) | [Go to Incremental Search](#)

Phylum	Class	Species	tRNA Seq. Anticodon
Bacteria	Actinobacteria	(3)	tRNA Seq. Anticodon
	Alphaproteobacteria	(116)	tRNA Seq. Anticodon
	Aquifae	(5)	tRNA Seq. Anticodon
	Bacteroidetes	(18)	tRNA Seq. Anticodon
	Betaproteobacteria	(70)	tRNA Seq. Anticodon
	Chlamydiae	(15)	tRNA Seq. Anticodon
	Chlorobi	(11)	tRNA Seq. Anticodon
	Chloroflexi	(10)	tRNA Seq. Anticodon

Anticodon table

		Number of tRNA genes							
		C	A	G	T				
T	AAA	0	ACG	2	Tyr				
	GAA	1230	GGA	982	GTA	1375			
C	AAU	1122	TGA	1224	Ser	TTA	0		
	CAA	880	CGA	652	Amber	CTA	1		
G	AAG	93	AGG	2	ATG	0	ACG	1615	
	GAG	850	GCG	626	GTC	1061	Arg	216	
A	ATG	1154	TGG	1236	Gln	TTG	1388		
	ATC	1009	CGG	503	CTG	466	Cys	766	
U	ATG	22	AGT	22	Asn	ATT	0	ACT	0
	ATC	2105	AGT	982	GTT	1633	Ser	992	
Ile	ATG	7	TGT	1354	Lys	TTT	1095	TCT	1039
	ATC	4016	CGT	596	CTT	590	Arg	731	
Val	AAU	0	AGC	0	ATC	0	ACC	0	
	GAC	931	GCG	882	GTC	1063	Gly	2020	
Ser	TAC	1913	TAC	2241	Glu	TTC	2149	TCC	1244
	TAC	345	CGC	323	CTG	351	Cys	522	

(b) tRNA gene list

tRNA gene List

Primates: *Macaca satellit* subsp. *satellit* str. 188
 MT-188 (segment)

Search: [tRNA gene list](#) | [Composite table](#) | [Composite table](#) | [Composite table](#)

(c) Advanced Pattern Search against each stem/loop

Advanced Pattern Search

Result: 1. Open reading frame (ORF) sequence
 2. Acceptor stem position 1-19
 3. Acceptor stem position 2-19
 4. Acceptor stem position 3-19
 5. Acceptor stem position 4-19
 6. Acceptor stem position 5-19
 7. Acceptor stem position 6-19
 8. Acceptor stem position 7-19
 9. Acceptor stem position 8-19
 10. Acceptor stem position 9-19
 11. Acceptor stem position 10-19
 12. Acceptor stem position 11-19
 13. Acceptor stem position 12-19
 14. Acceptor stem position 13-19
 15. Acceptor stem position 14-19
 16. Acceptor stem position 15-19
 17. Acceptor stem position 16-19
 18. Acceptor stem position 17-19
 19. Acceptor stem position 18-19
 20. Acceptor stem position 19-19
 21. Acceptor stem position 20-19
 22. Acceptor stem position 21-19
 23. Acceptor stem position 22-19
 24. Acceptor stem position 23-19
 25. Acceptor stem position 24-19
 26. Acceptor stem position 25-19
 27. Acceptor stem position 26-19
 28. Acceptor stem position 27-19
 29. Acceptor stem position 28-19
 30. Acceptor stem position 29-19
 31. Acceptor stem position 30-19
 32. Acceptor stem position 31-19
 33. Acceptor stem position 32-19
 34. Acceptor stem position 33-19
 35. Acceptor stem position 34-19
 36. Acceptor stem position 35-19
 37. Acceptor stem position 36-19
 38. Acceptor stem position 37-19
 39. Acceptor stem position 38-19
 40. Acceptor stem position 39-19
 41. Acceptor stem position 40-19
 42. Acceptor stem position 41-19
 43. Acceptor stem position 42-19
 44. Acceptor stem position 43-19
 45. Acceptor stem position 44-19
 46. Acceptor stem position 45-19
 47. Acceptor stem position 46-19
 48. Acceptor stem position 47-19
 49. Acceptor stem position 48-19
 50. Acceptor stem position 49-19
 51. Acceptor stem position 50-19
 52. Acceptor stem position 51-19
 53. Acceptor stem position 52-19
 54. Acceptor stem position 53-19
 55. Acceptor stem position 54-19
 56. Acceptor stem position 55-19
 57. Acceptor stem position 56-19
 58. Acceptor stem position 57-19
 59. Acceptor stem position 58-19
 60. Acceptor stem position 59-19
 61. Acceptor stem position 60-19
 62. Acceptor stem position 61-19
 63. Acceptor stem position 62-19
 64. Acceptor stem position 63-19
 65. Acceptor stem position 64-19
 66. Acceptor stem position 65-19
 67. Acceptor stem position 66-19
 68. Acceptor stem position 67-19
 69. Acceptor stem position 68-19
 70. Acceptor stem position 69-19
 71. Acceptor stem position 70-19
 72. Acceptor stem position 71-19
 73. Acceptor stem position 72-19
 74. Acceptor stem position 73-19
 75. Acceptor stem position 74-19
 76. Acceptor stem position 75-19
 77. Acceptor stem position 76-19
 78. Acceptor stem position 77-19
 79. Acceptor stem position 78-19
 80. Acceptor stem position 79-19
 81. Acceptor stem position 80-19
 82. Acceptor stem position 81-19
 83. Acceptor stem position 82-19
 84. Acceptor stem position 83-19
 85. Acceptor stem position 84-19
 86. Acceptor stem position 85-19
 87. Acceptor stem position 86-19
 88. Acceptor stem position 87-19
 89. Acceptor stem position 88-19
 90. Acceptor stem position 89-19
 91. Acceptor stem position 90-19
 92. Acceptor stem position 91-19
 93. Acceptor stem position 92-19
 94. Acceptor stem position 93-19
 95. Acceptor stem position 94-19
 96. Acceptor stem position 95-19
 97. Acceptor stem position 96-19
 98. Acceptor stem position 97-19
 99. Acceptor stem position 98-19
 100. Acceptor stem position 99-19
 101. Acceptor stem position 100-19
 102. Acceptor stem position 101-19
 103. Acceptor stem position 102-19
 104. Acceptor stem position 103-19
 105. Acceptor stem position 104-19
 106. Acceptor stem position 105-19
 107. Acceptor stem position 106-19
 108. Acceptor stem position 107-19
 109. Acceptor stem position 108-19
 110. Acceptor stem position 109-19
 111. Acceptor stem position 110-19
 112. Acceptor stem position 111-19
 113. Acceptor stem position 112-19
 114. Acceptor stem position 113-19
 115. Acceptor stem position 114-19
 116. Acceptor stem position 115-19
 117. Acceptor stem position 116-19
 118. Acceptor stem position 117-19
 119. Acceptor stem position 118-19
 120. Acceptor stem position 119-19
 121. Acceptor stem position 120-19
 122. Acceptor stem position 121-19
 123. Acceptor stem position 122-19
 124. Acceptor stem position 123-19
 125. Acceptor stem position 124-19
 126. Acceptor stem position 125-19
 127. Acceptor stem position 126-19
 128. Acceptor stem position 127-19
 129. Acceptor stem position 128-19
 130. Acceptor stem position 129-19
 131. Acceptor stem position 130-19
 132. Acceptor stem position 131-19
 133. Acceptor stem position 132-19
 134. Acceptor stem position 133-19
 135. Acceptor stem position 134-19
 136. Acceptor stem position 135-19
 137. Acceptor stem position 136-19
 138. Acceptor stem position 137-19
 139. Acceptor stem position 138-19
 140. Acceptor stem position 139-19
 141. Acceptor stem position 140-19
 142. Acceptor stem position 141-19
 143. Acceptor stem position 142-19
 144. Acceptor stem position 143-19
 145. Acceptor stem position 144-19
 146. Acceptor stem position 145-19
 147. Acceptor stem position 146-19
 148. Acceptor stem position 147-19
 149. Acceptor stem position 148-19
 150. Acceptor stem position 149-19
 151. Acceptor stem position 150-19
 152. Acceptor stem position 151-19
 153. Acceptor stem position 152-19
 154. Acceptor stem position 153-19
 155. Acceptor stem position 154-19
 156. Acceptor stem position 155-19
 157. Acceptor stem position 156-19
 158. Acceptor stem position 157-19
 159. Acceptor stem position 158-19
 160. Acceptor stem position 159-19
 161. Acceptor stem position 160-19
 162. Acceptor stem position 161-19
 163. Acceptor stem position 162-19
 164. Acceptor stem position 163-19
 165. Acceptor stem position 164-19
 166. Acceptor stem position 165-19
 167. Acceptor stem position 166-19
 168. Acceptor stem position 167-19
 169. Acceptor stem position 168-19
 170. Acceptor stem position 169-19
 171. Acceptor stem position 170-19
 172. Acceptor stem position 171-19
 173. Acceptor stem position 172-19
 174. Acceptor stem position 173-19
 175. Acceptor stem position 174-19
 176. Acceptor stem position 175-19
 177. Acceptor stem position 176-19
 178. Acceptor stem position 177-19
 179. Acceptor stem position 178-19
 180. Acceptor stem position 179-19
 181. Acceptor stem position 180-19
 182. Acceptor stem position 181-19
 183. Acceptor stem position 182-19
 184. Acceptor stem position 183-19
 185. Acceptor stem position 184-19
 186. Acceptor stem position 185-19
 187. Acceptor stem position 186-19
 188. Acceptor stem position 187-19
 189. Acceptor stem position 188-19
 190. Acceptor stem position 189-19
 191. Acceptor stem position 190-19
 192. Acceptor stem position 191-19
 193. Acceptor stem position 192-19
 194. Acceptor stem position 193-19
 195. Acceptor stem position 194-19
 196. Acceptor stem position 195-19
 197. Acceptor stem position 196-19
 198. Acceptor stem position 197-19
 199. Acceptor stem position 198-19
 200. Acceptor stem position 199-19
 201. Acceptor stem position 200-19
 202. Acceptor stem position 201-19
 203. Acceptor stem position 202-19
 204. Acceptor stem position 203-19
 205. Acceptor stem position 204-19
 206. Acceptor stem position 205-19
 207. Acceptor stem position 206-19
 208. Acceptor stem position 207-19
 209. Acceptor stem position 208-19
 210. Acceptor stem position 209-19
 211. Acceptor stem position 210-19
 212. Acceptor stem position 211-19
 213. Acceptor stem position 212-19
 214. Acceptor stem position 213-19
 215. Acceptor stem position 214-19
 216. Acceptor stem position 215-19
 217. Acceptor stem position 216-19
 218. Acceptor stem position 217-19
 219. Acceptor stem position 218-19
 220. Acceptor stem position 219-19
 221. Acceptor stem position 220-19
 222. Acceptor stem position 221-19
 223. Acceptor stem position 222-19
 224. Acceptor stem position 223-19
 225. Acceptor stem position 224-19
 226. Acceptor stem position 225-19
 227. Acceptor stem position 226-19
 228. Acceptor stem position 227-19
 229. Acceptor stem position 228-19
 230. Acceptor stem position 229-19
 231. Acceptor stem position 230-19
 232. Acceptor stem position 231-19
 233. Acceptor stem position 232-19
 234. Acceptor stem position 233-19
 235. Acceptor stem position 234-19
 236. Acceptor stem position 235-19
 237. Acceptor stem position 236-19
 238. Acceptor stem position 237-19
 239. Acceptor stem position 238-19
 240. Acceptor stem position 239-19
 241. Acceptor stem position 240-19
 242. Acceptor stem position 241-19
 243. Acceptor stem position 242-19
 244. Acceptor stem position 243-19
 245. Acceptor stem position 244-19
 246. Acceptor stem position 245-19
 247. Acceptor stem position 246-19
 248. Acceptor stem position 247-19
 249. Acceptor stem position 248-19
 250. Acceptor stem position 249-19
 251. Acceptor stem position 250-19
 252. Acceptor stem position 251-19
 253. Acceptor stem position 252-19
 254. Acceptor stem position 253-19
 255. Acceptor stem position 254-19
 256. Acceptor stem position 255-19
 257. Acceptor stem position 256-19
 258. Acceptor stem position 257-19
 259. Acceptor stem position 258-19
 260. Acceptor stem position 259-19
 261. Acceptor stem position 260-19
 262. Acceptor stem position 261-19
 263. Acceptor stem position 262-19
 264. Acceptor stem position 263-19
 265. Acceptor stem position 264-19
 266. Acceptor stem position 265-19
 267. Acceptor stem position 266-19
 268. Acceptor stem position 267-19
 269. Acceptor stem position 268-19
 270. Acceptor stem position 269-19
 271. Acceptor stem position 270-19
 272. Acceptor stem position 271-19
 273. Acceptor stem position 272-19
 274. Acceptor stem position 273-19
 275. Acceptor stem position 274-19
 276. Acceptor stem position 275-19
 277. Acceptor stem position 276-19
 278. Acceptor stem position 277-19
 279. Acceptor stem position 278-19
 280. Acceptor stem position 279-19
 281. Acceptor stem position 280-19
 282. Acceptor stem position 281-19
 283. Acceptor stem position 282-19
 284. Acceptor stem position 283-19
 285. Acceptor stem position 284-19
 286. Acceptor stem position 285-19
 287. Acceptor stem position 286-19
 288. Acceptor stem position 287-19
 289. Acceptor stem position 288-19
 290. Acceptor stem position 289-19
 291. Acceptor stem position 290-19
 292. Acceptor stem position 291-19
 293. Acceptor stem position 292-19
 294. Acceptor stem position 293-19
 295. Acceptor stem position 294-19
 296. Acceptor stem position 295-19
 297. Acceptor stem position 296-19
 298. Acceptor stem position 297-19
 299. Acceptor stem position 298-19
 300. Acceptor stem position 299-19
 301. Acceptor stem position 300-19
 302. Acceptor stem position 301-19
 303. Acceptor stem position 302-19
 304. Acceptor stem position 303-19
 305. Acceptor stem position 304-19
 306. Acceptor stem position 305-19
 307. Acceptor stem position 306-19
 308. Acceptor stem position 307-19
 309. Acceptor stem position 308-19
 310. Acceptor stem position 309-19
 311. Acceptor stem position 310-19
 312. Acceptor stem position 311-19
 313. Acceptor stem position 312-19
 314. Acceptor stem position 313-19
 315. Acceptor stem position 314-19
 316. Acceptor stem position 315-19
 317. Acceptor stem position 316-19
 318. Acceptor stem position 317-19
 319. Acceptor stem position 318-19
 320. Acceptor stem position 319-19
 321. Acceptor stem position 320-19
 322. Acceptor stem position 321-19
 323. Acceptor stem position 322-19
 324. Acceptor stem position 323-19
 325. Acceptor stem position 324-19
 326. Acceptor stem position 325-19
 327. Acceptor stem position 326-19
 328. Acceptor stem position 327-19
 329. Acceptor stem position 328-19
 330. Acceptor stem position 329-19
 331. Acceptor stem position 330-19
 332. Acceptor stem position 331-19
 333. Acceptor stem position 332-19
 334. Acceptor stem position 333-19
 335. Acceptor stem position 334-19
 336. Acceptor stem position 335-19
 337. Acceptor stem position 336-19
 338. Acceptor stem position 337-19
 339. Acceptor stem position 338-19
 340. Acceptor stem position 339-19
 341. Acceptor stem position 340-19
 342. Acceptor stem position 341-19
 343. Acceptor stem position 342-19
 344. Acceptor stem position 343-19
 345. Acceptor stem position 344-19
 346. Acceptor stem position 345-19
 347. Acceptor stem position 346-19
 348. Acceptor stem position 347-19
 349. Acceptor stem position 348-19
 350. Acceptor stem position 349-19
 351. Acceptor stem position 350-19
 352. Acceptor stem position 351-19
 353. Acceptor stem position 352-19
 354. Acceptor stem position 353-19
 355. Acceptor stem position 354-19
 356. Acceptor stem position 355-19
 357. Acceptor stem position 356-19
 358. Acceptor stem position 357-19
 359. Acceptor stem position 358-19
 360. Acceptor stem position 359-19
 361. Acceptor stem position 360-19
 362. Acceptor stem position 361-19
 363. Acceptor stem position 362-19
 364. Acceptor stem position 363-19
 365. Acceptor stem position 364-19
 366. Acceptor stem position 365-19
 367. Acceptor stem position 366-19
 368. Acceptor stem position 367-19
 369. Acceptor stem position 368-19
 370. Acceptor stem position 369-19
 371. Acceptor stem position 370-19
 372. Acceptor stem position 371-19
 373. Acceptor stem position 372-19
 374. Acceptor stem position 373-19
 375. Acceptor stem position 374-19
 376. Acceptor stem position 375-19
 377. Acceptor stem position 376-19
 378. Acceptor stem position 377-19
 379. Acceptor stem position 378-19
 380. Acceptor stem position 379-19
 381. Acceptor stem position 380-19
 382. Acceptor stem position 381-19
 383. Acceptor stem position 382-19
 384. Acceptor stem position 383-19
 385. Acceptor stem position 384-19
 386. Acceptor stem position 385-19
 387. Acceptor stem position 386-19
 388. Acceptor stem position 387-19
 389. Acceptor stem position 388-19
 390. Acceptor stem position 389-19
 391. Acceptor stem position 390-19
 392. Acceptor stem position 391-19
 393. Acceptor stem position 392-19
 394. Acceptor stem position 393-19
 395. Acceptor stem position 394-19
 396. Acceptor stem position 395-19
 397. Acceptor stem position 396-19
 398. Acceptor stem position 397-19
 399. Acceptor stem position 398-19
 400. Acceptor stem position 399-19
 401. Acceptor stem position 400-19
 402. Acceptor stem position 401-19
 403. Acceptor stem position 402-19
 404. Acceptor stem position 403-19
 405. Acceptor stem position 404-19
 406. Acceptor stem position 405-19
 407. Acceptor stem position 406-19
 408. Acceptor stem position 407-19
 409. Acceptor stem position 408-19
 410. Acceptor stem position 409-19
 411. Acceptor stem position 410-19
 412. Acceptor stem position 411-19
 413. Acceptor stem position 412-19
 414. Acceptor stem position 413-19
 415. Acceptor stem position 414-19
 416. Acceptor stem position 415-19
 417. Acceptor stem position 416-19
 418. Acceptor stem position 417-19
 419. Acceptor stem position 418-19
 420. Acceptor stem position 419-19
 421. Acceptor stem position 420-19
 422. Acceptor stem position 421-19
 423. Acceptor stem position 422-19
 424. Acceptor stem position 423-19
 425. Acceptor stem position 424-19
 426. Acceptor stem position 425-19
 427. Acceptor stem position 426-19
 428. Acceptor stem position 427-19
 429. Acceptor stem position 428-19
 430. Acceptor stem position 429-19
 431. Acceptor stem position 430-19
 432. Acceptor stem position 431-19
 433. Acceptor stem position 432-19
 434. Acceptor stem position 433-19
 435. Acceptor stem position 434-19
 436. Acceptor stem position 435-1

either the reliable tRNA genes or all candidate genes by choosing ‘The reliable tRNA genes’ or ‘All candidate genes’. Criteria used for this manual evaluation were described in detail previously (4). The tRNA genes of Archaea obtained from SPLITSdb (8) were included in this new version. Basic functions of the database were described previously in detail and briefly in Figure 1.

To aim at establishing a very reliable database utilizing collective knowledge in the various experimental fields of tRNA research, we developed a new function for including comments on each of the registered tRNA genes on ‘the detailed information of tRNA gene sequence page’ (Figure 1d). User can add comments by entering e-mail address and password, while we reserve the right to remove irrelevant comments. We hope that the accumulation of user comments will provide annotation of higher quality and that this database will become an information sharing system in the tRNA gene community.

Identical sequence groups and their use as phylogenetic markers for environmental metagenomic sequences

When we conducted the clustering of 124 015 sequences of tRNA genes, except for the 3'-CCA terminal sequence, from prokaryotic genomes by sequence alignment using the CD-HIT (9), we found high phylogenetic preservation of tRNA genes: i.e. a particular tRNA sequence was found only in a particular lineage of phylogenetic groups. We designated here the tRNA group with an identical sequence as ‘Identical Sequence Group: ISG’ (Figure 2a) and listed the numbers of ISGs for each phylotype (Figure 2b) and for each anticodon (Figure 2c). The tRNAs with one anticodon type were classified and listed according to ISG along with the phylotype information of each tRNA (Figure 2d), and therefore, the range of phylotypes found for each ISG could be examined. If we focused on ISGs composed of more than five sequences, 97.1% of ISGs were conserved at a phylum level, showing most tRNAs to be good phylogenetic markers at least at a phylum level. The ISGs could provide a strategy

(a) Identical Sequence Group (ISG)

Data Type		Genomes	Sequence groups
Prokaryote	Bacteria	2181	28061
	Archaea	47	1690
Eukaryote	Plant	2	438
	Fungi	10	511
Virus		31	97
Phage		120	486
Chloroplast		121	1731
ENV		123	34387
ENV from Sequence Read Archive (SRA)		87	26188

Identical Sequence Group (ENV vs Prokaryote Genome)

100% (Identical) | 97% (2-nt difference) | 95% (3-nt difference)

Identical Sequence Group (SRA vs Prokaryote Genome)

100% (Identical) | 97% (2-nt difference) | 95% (3-nt difference)

(b) ISGs for each phylotype

Phylum/Class/Species List	
Search species :	[Search]
Bacteria(2181)	Identical Group(28061)
Acidobacteria (4)	Identical Group(134)
Actinobacteria (197)	Identical Group(3501)
Alphaproteobacteria (253)	Identical Group(3242)
Aquificeae (9)	Identical Group(236)
Bacteroidetes (85)	Identical Group(1814)
Betaproteobacteria (183)	Identical Group(1463)
Chlamydiae (22)	Identical Group(152)
Chlorobi (18)	Identical Group(300)

(d) ISG tRNAs for one anticodon

tRNA genes clustered and arranged according to each identical group									
Anticodon = GAA HIT: 565 Group(s)									
Download Sequence [FASTA] [TAB] [Representative sequence of each cluster] View ClusterW result [ClusterW]									
<input checked="" type="checkbox"/> Select [check all]									
Sequence ID (or Accession No.)									
Genome ID (Sample source for ENV)									
Phylum/Class (Sample source for ENV)									
Species									
Start									
End									
Direction									
AA									
Anticodon									
Genome/Sq. Info.									
Decision									

(c) ISGs for each anticodon

Identical Sequence Group									
Bacteria									
Bacteria(Complete & Draft)									
100 %									
>tRNA [ISG] [Identical Group] [ISG within this phylotype] [ISG outside this phylotype]									
Number of tRNA clusters									
Subtable									
T	phe	AAA	1	AGA	2	Tyr	ATA	D	Cys
Lys	GAA	843	GGG	738	748	Arg	GCA	0	T
Leu	GAA	867	TCA	814	746	Isole	TAA	1	Arg Sec
C	AAU	90	AGG	2	ATG	1	Arg	602	T
Ile	UAG	668	GCG	426	TGA	745	GCG	64	C
Leu	UAG	827	TGA	602	TTC	792	TGC	227	A
C	AAU	634	CGG	377	CTG	363	CCG	600	G
A	UAT	2	AGT	19	ATT	—	ACT	1	T
Ile	UAT	521	Tyr	602	GTT	720	GCT	826	C
Met	CAT	2926	CGT	519	CTT	470	CCT	712	G
G	AAU	2	AGC	0	ATC	0	ACC	0	T
Val	GAC	521	GCG	437	GTC	603	GCC	544	C
		674	TGC	500	TTC	654	TCC	571	A
		282	CGC	268	GTC	329	CCC	422	G

(d) ISG tRNAs for one anticodon

tRNA genes clustered and arranged according to each identical group									
Anticodon = GAA HIT: 565 Group(s)									
Download Sequence [FASTA] [TAB] [Representative sequence of each cluster] View ClusterW result [ClusterW]									
<input checked="" type="checkbox"/> Select [check all]									
Sequence ID (or Accession No.)									
Genome ID (Sample source for ENV)									
Phylum/Class (Sample source for ENV)									
Species									
Start									
End									
Direction									
AA									
Anticodon									
Genome/Sq. Info.									
Decision									

(e) ISGs found in an environmental sample

Sample List	
ENV (12)	
Acid Mine Drainage microbial community from Richmond mine	(1)
Complete metagenome contig#0001	(1)
Corall interponeme fab_M02_001	(1)
Corall interponeme fab_M02_002	(1)
Corall interponeme: microbial fraction from Polycarpian corallumensis from coral tissue contains sequences from fungi (presumably yeasts), bacteria, viruses, and DNA	(1)
coral mitochondrial	(1)
Environmental sample from EWV division of INSOC	(1)
Fossil microbial community from Whale Fall (the West Antarctic Peninsula Shelf)	(1)
Fossil microbial community from whale fall (West Antarctic Peninsula Shelf)	(1)
Fossil microbial from Benthos prasinorum	(1)
Fossil microbial from Neanderthal	(1)
Freshwater sediment metagenome Informaticle_C1	(1)
Freshwater sediment metagenome Informaticle_C1	(1)

tRNA genes obtained from an environmental sample assignable to identical sequence group.									
HIT: 153 sequence(s).									
Download Sequence [FASTA] [TAB]									
View ClusterW result [ClusterW]									
<input checked="" type="checkbox"/> Relative assignment <input type="checkbox"/> Tetralic assignment									
Select [check all]									
Sequence ID (or Accession No.)									
Genome ID (or Accession No.)									
Identical group No.									
Phylum/Class (Sample source for ENV)									
Anticodon									
AA									

tRNA genes obtained from an environmental sample assignable to identical sequence group.									
HIT: 153 sequence(s).									
Download Sequence [FASTA] [TAB]									
View ClusterW result [ClusterW]									
<input checked="" type="checkbox"/> Relative assignment <input type="checkbox"/> Tetralic assignment									
Select [check all]									
Sequence ID (or Accession No.)									
Genome ID (or Accession No.)									
Identical group No.									
Phylum/Class (Sample source for ENV)									
Anticodon									
AA									

to select reliable phylogenetic markers. At the genus level, ~60% of ISGs were conserved, showing there may exist good genus-specific markers. By combining the data provided by this database with other detail knowledge of a particular tRNA obtained by experiments or from literature, users may get useful phylogenetic markers (e.g. genus-specific markers) by themselves.

Interestingly, among 154 455 tRNA genes found in metagenomic sequences derived from environmental samples, 35 739 tRNA genes (23%) were identical in sequence to genes from known prokaryotes. Using tRNAs found in an environment sample that were assigned to ISGs, we could predict microbial community structures in an environmental ecosystem at least at a phylum level (Figure 2e). The database has also a function to search for sequences with 97 or 95% sequence identify (2- or 3-nt difference, respectively) (Figure 2a). By using tools in the database and the specific markers found by users (e.g. genus-specific markers), users can independently clarify microbial populations in an ecosystem. This strategy can be applied even to data of short sequences obtained using next-generation sequencers, such as SRA in NCBI. In metagenomic analysis by a next-generation sequencer where the length of sequences obtained is short, phylogenetic characterization of the short sequences was particularly difficult using existing methods, except for sequences derived from dominant species or sequences unambiguously mapped on a known sequenced genome. Because tRNA genes are searchable even from short genomic fragments of around 100 bases, tRNA genes should become one of the most effective means for identifying the microbial populations in an ecosystem when analyzing a vast number of metagenomic sequences obtained by next-generation sequencers.

ACCESS TO THE DATABASE

tRNADB-CE can be accessed freely from <http://trna.nagahama-i-bio.ac.jp>.

FUNDING

Integrated Database Project and a Grant-in-Aid for Publication of Scientific Research Results from the Ministry of Education, Culture, Sports, Science and Technology, Japan. Funding for open access charge: Integrated Database Project from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Conflict of interest statement. None declared.

REFERENCES

1. Amann,R.I., Ludwig,W. and Schleifer,K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.*, **59**, 143–169.
2. Delong,E.F. and Karl,D.M. (2005) Genomic perspectives in microbial oceanography. *Nature*, **437**, 336–342.
3. Denef,V.J., Mueller,R.S. and Banfield,J.F. (2010) AMD biofilms: using model communities to study evolution and ecological complexity in nature. *ISME J.*, **4**, 599–610.
4. Abe,T., Ikemura,T., Ohara,Y., Uehara,H., Kinouchi,M., Kanaya,S., Yamada,Y., Muto,A. and Inokuchi,H. (2009) tRNADB-CE: tRNA gene database curated manually by experts. *Nucleic Acids Res.*, **37**, D163–D168.
5. Lowe,T.M. and Eddy,S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, **25**, 955–964.
6. Laslett,D. and Canback,B. (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.*, **32**, 11–16.
7. Kinouchi,M. and Kurokawa,K. (2006) tRNAdesigner: A software system to find all tRNA genes in the DNA sequence based on the cloverleaf secondary structure. *J. Comput. Aided Chem.*, **7**, 116–126.
8. Sugahara,J., Kikuta,K., Fujishima,K., Yachie,N., Tomita,M. and Kanai,A. (2008) Comprehensive analysis of archaeal tRNA genes reveals rapid increase of tRNA introns in the order thermoproteales. *Mol. Biol. Evol.*, **25**, 2709–2716.
9. Li,W. and Godzik,A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.